Comparisons of SARS-CoV-2 serological surveillance across multiple data sources

Yuan Yu1, Sheila O’Brien2, David L. Buckeridge1, W. Alton Russell1

1School of Population and Global Health, McGill University, Montreal, Canada

2 Canadian Blood Services, Ottawa, Canada

**Corresponding author:**

**Key words:**

**Running title:**

##### 

# Abstract

Blood donors serve as a crucial population for serosurveillance due to its convenient data collection process compared to other public health surveys. The fact that it consists large proportion of the elder male population might throw doubts on the seropositivity estimation for the general population. In this paper, firstly, we compared the seropositivity estimation of the general population across different data resources for varies time period, evaluating the effect of multilevel regression and poststratification (MRP) in comparison to traditional statistical adjustment approach like raking. Secondly, we further tracked the temporal pattern of the MRP adjusted seropositivity curve to find that representation error is not the main cause of the serosurveillance discrepancy across different data resources, by evaluating the average absolute difference and proportion of days of closer estimations before and after the evaluation. The gap of serosurveillance for the general population among different resources depend on varies factors including the data collection, blood sample type (serum or dried blood spots), assay difference and operating labs.

**Background:** A

**Methods:** A

**Results:** A

**Conclusions:** A

##### 

# Introduction

* There are multiple serosurveillance studies through tracking the anti-N seropositivity across different populations.
* Brief description for each data source and the different nature of each dataset.
* Brief intro of the conventional raking approach, which can be done based on demographics, geography, and neighborhood-level characteristics. MRP is a classic approach to adjust for the representative bias, earlier combination studies focused more on trend of the seroprevalance for combined studies while exploring the effect of different representativeness across different studies and ideal methodology of suitable adjustment is our main purpose

# Methods

**Data**: CBS, CCAHS, ABC, APL, CLSA, CanPath

**Modeling**: Multilevel Regression and Poststritification (Theorem, regression variables, poststratification table)

1. MRP
2. Temporal MRP

# Results

The MRP generally provides more precise estimates with lower CV (coefficient of variance = mean/sd) and admissible compared with other adjusting methods like raking and raw post-stratification, especially adjusting for multiple demographic features when aggregation table is sparse for smaller sample size (i.e. AbC).

Especially, for later Omicron period, MRP could pull the seropositivity estimation together in comparing the adjusted and unadjusted seropositivity comparing the CBS and CCAHS II, for both overall and temporal seropositivity estimation. However, MRP does not seem to have the same effect on the earlier period when comparing with other data resources, no matter for overall or temporal seropositivity.

Even within the same data source, MRP does not adjust the seropositivity in the same direction. For example, it tends to adjust the trend higher for the Prairies regions during the earlier Omicron but lower for the latter period.

Since there is still an obvious gap between the population level seropositivity curves after adjusted for the potential representation error. For evaluating the potential cause of the discrepancy, we provide the coefficient analysis of the log-odds for a combined data source including the type of the blood. It showed that the higher seropositivity might be caused by serum blood samples collected by the CBS, compared to the dried blood spots used in AbC.

# Discussion

* Further thoughts and explanations of the results:

1. Discuss the results for different settings (time cuttings).
2. Discuss the results from temporal MRP and the advantage of adding the time component.
3. List the advantages of the MRP in comparison to the traditional statistical adjustment like raking:

a) Automatically able to adjust for more variables while raking would fail for lacking of the marginal counts especially when sample size is small or trying to adjust for more demographic variables

b) From the performance of the estimation, the CV (coefficient of the variance) tend to be smaller, but not always hold for true.

1. However, MRP only pull the estimations for particular time period among specific comparison datasets, CBS and CCAHS in our case.

* Limitations:

1. Different anti-seropositivity measured by different assays on different type of blood samples. Anti-N seropositivity is lack of universal method to convert to a consistent measure to reduce the potential bias induced by the assay difference.
2. All seropositivity estimates are Rogan-Gladen adjusted, which partially adjust for the population level seropositivity. However, it might impact the seropositivity estimation more for assay with low sensitivity.
3. Other potential bias: Collection date might not available, multiple definitions of seropositivity, unknown lab difference.
4. Surveys were largely not designed for estimating changing seropositivity over time.

* Conclusions:

Other than the representative bias, the gap of serosurveillance for the general population among different resources depend on varies factors including the data collection, blood sample type (serum or dried blood spots), assay difference and operating labs.

##### 

# Declarations

**Funding:** A

**Conflicts:** A

**Ethics/Consent:** A

**Data and materials:** A

**Code availability:** A

**Authors’ contributions:**

##### 

# References

1. Langham S, Wright A, Kenworthy J, Grieve R, Dunlop WCN. Cost-effectiveness of take-home naloxone for the prevention of overdose fatalities among heroin users in the United Kingdom. *Value in Health*. 2018;21(4):407-415. doi:[10.1016/j.jval.2017.07.014](https://doi.org/10.1016/j.jval.2017.07.014)

2. Keane C, Egan JE, Hawk M. Effects of naloxone distribution to likely bystanders: Results of an agent-based model. *International Journal of Drug Policy*. 2018;55:61-69. doi:[10.1016/j.drugpo.2018.02.008](https://doi.org/10.1016/j.drugpo.2018.02.008)

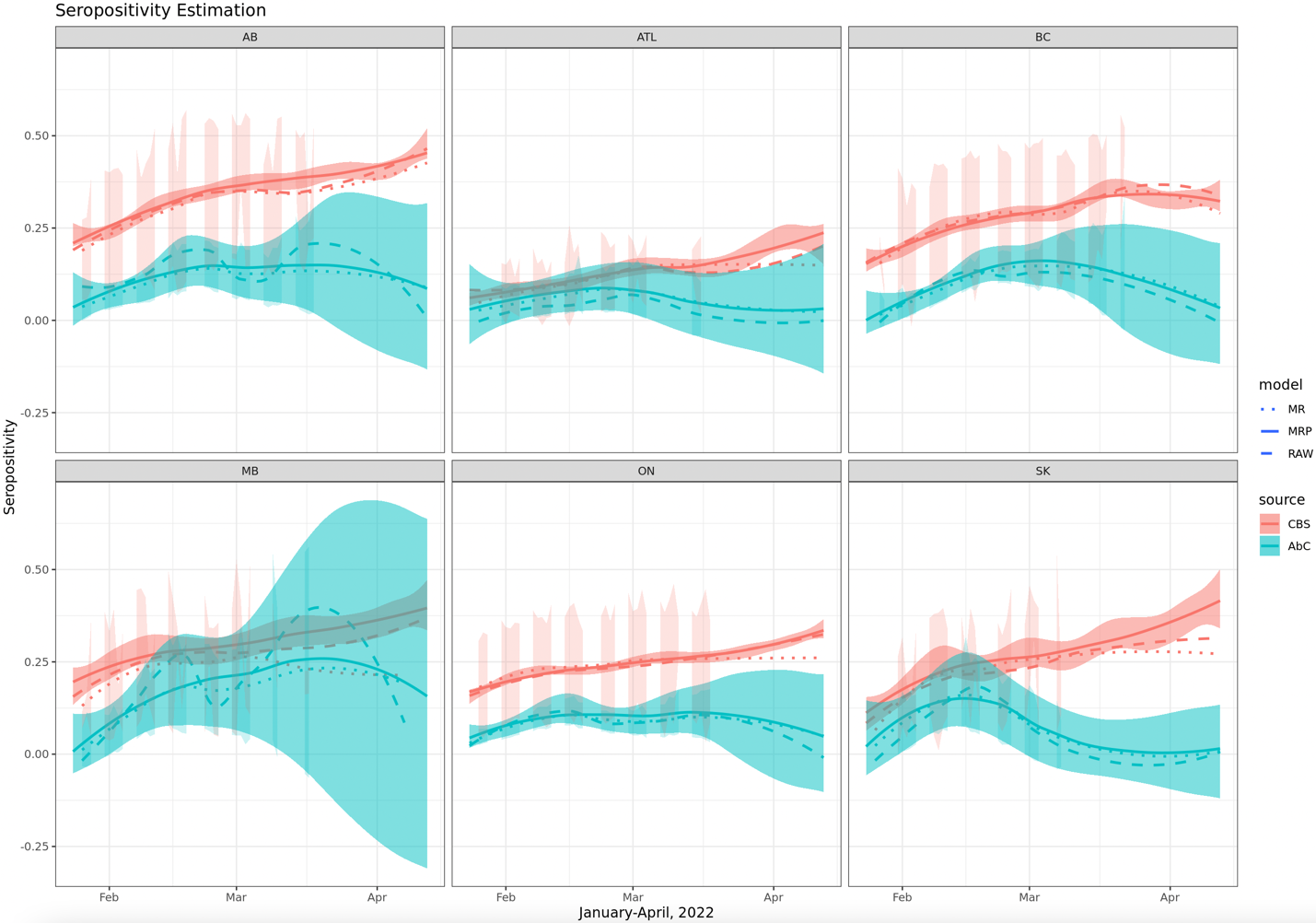
##### 

# Tables

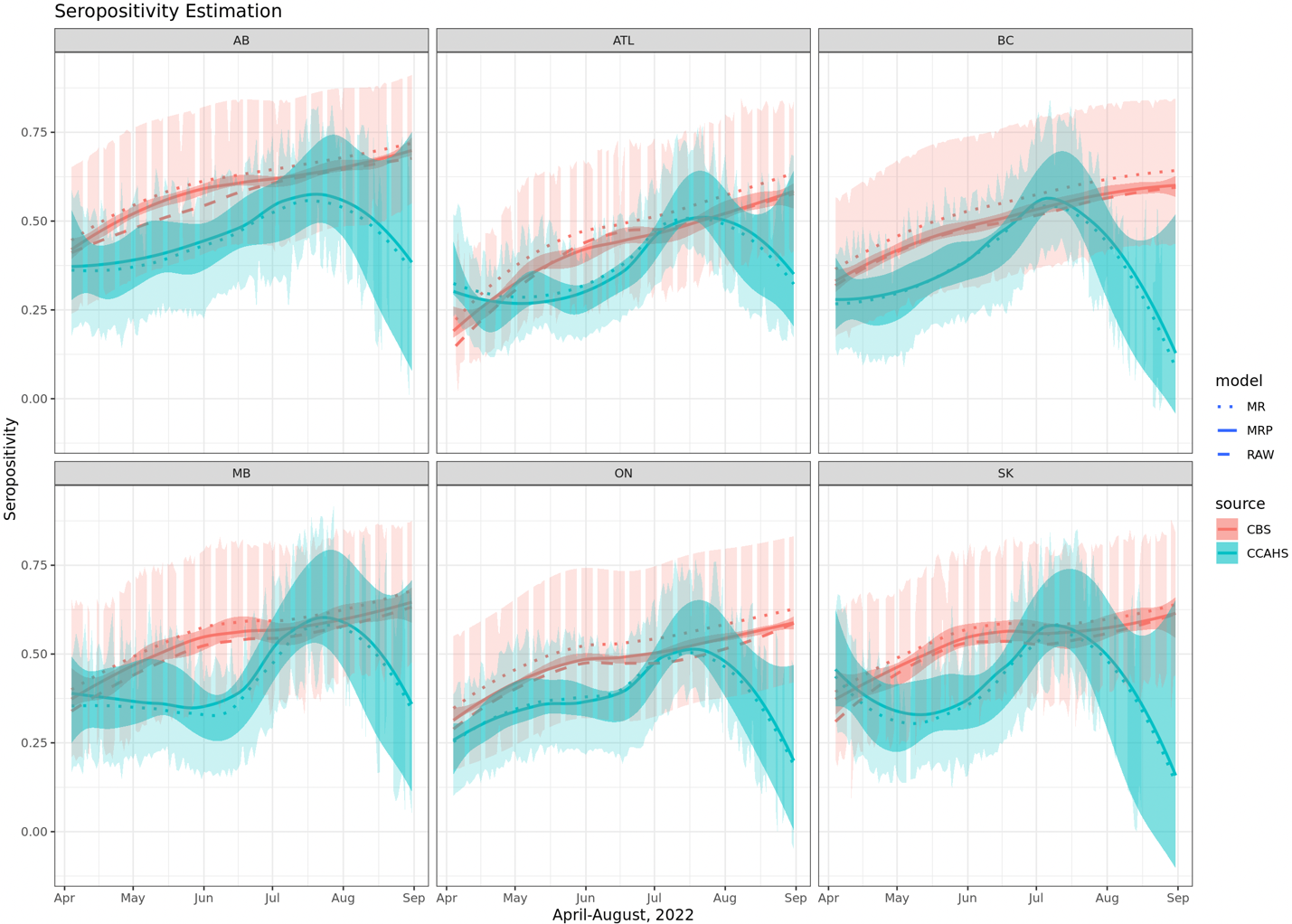
**Table 1:** Comparison of Average absolute difference (proportion of days of closer estimations of MRP) between CBS and other surveys for different settings (MRP adjustments for temporal estimates).

##### 

# Figures



**Figure 1:** Seropositivity estimation during Omicron, January - April, 2022. Average absolute difference between CBS vs AbC are 0.138 and 0.165 from MR and MRP model correspondingly; Proportion of days of closer estimations from MRP are 18%, 43%, 44%, 62%, 49%, 29% for Alberta, Atlantic region, British Columbia, Manitoba, Ontario and Saskatchewan. MRP does not make estimations from different sources closer for most overlapping time periods of CBS and AbC especially for AB and SK. Note: Darker shades indicate a loess smoothing for estimates of daily seroposotivity.



**Figure 2:** Seropositivity estimation during Omicron, April – August 2022. Average absolute difference between CBS vs CCAHS II are 0.141 and 0.1113 from MR and MRP model correspondingly; Proportion of days of closer estimations from MRP are 94%, 77%, 83%, 83%, 91%, 75% for Alberta, Atlantic region, British Columbia, Manitoba, Ontario and Saskatchewan. MRP makes estimations from different sources closer for most overlapping time periods of CBS and CCAHS II.

##### 

# Supplemental materials

# A. Supplement section

##### 

# Supplemental tables

**Table S1:** tableone for AbC.

**Table S2:** tableone for CanPath.

**Table S3:** tableone for CanPath.

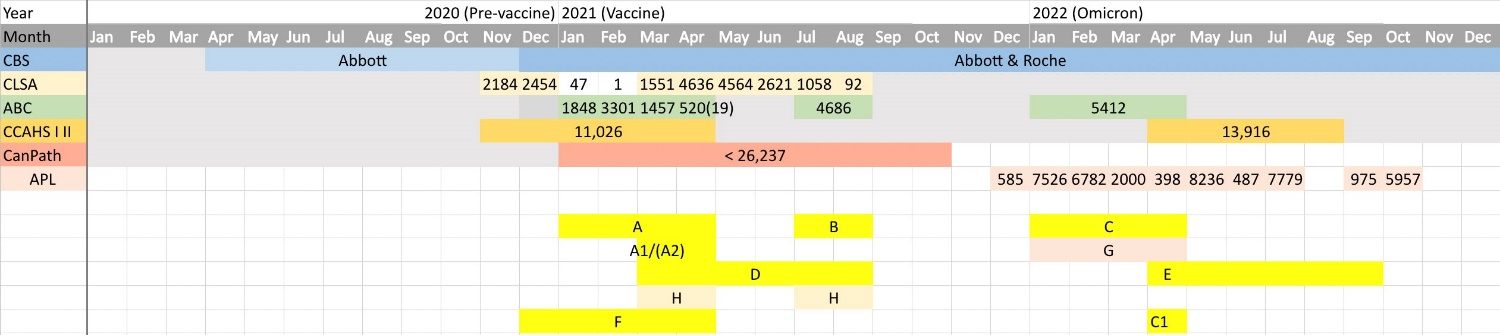
**Table S4:** tableone for CLSA.

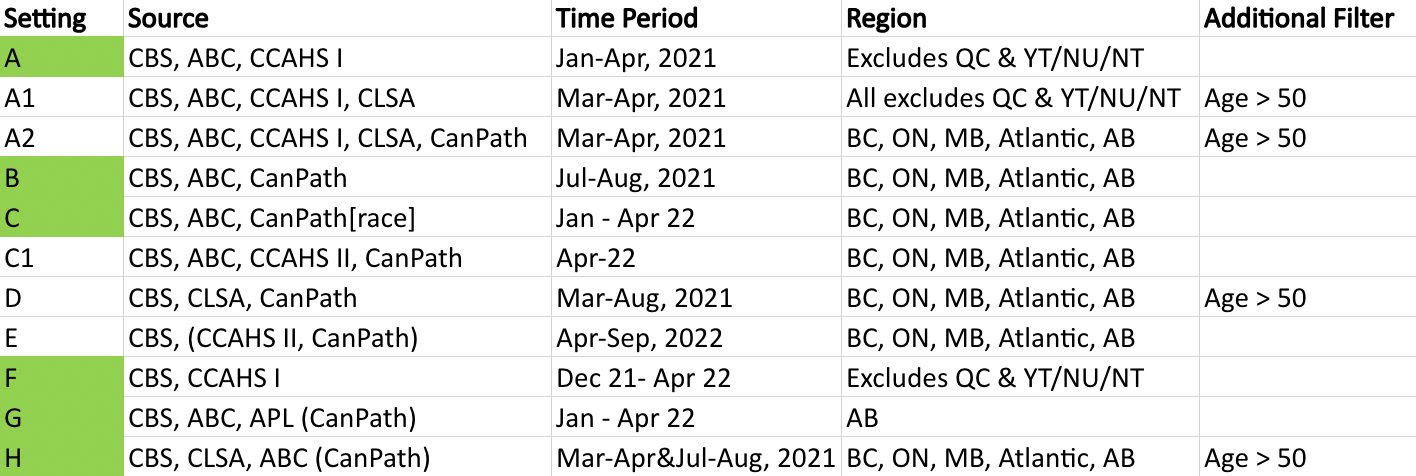
**Table S5:** tableone for APL.

**Table S6:** tableone for CCAHS I.

**Table S7:** tableone for CCAHS II.

**Table S7:** Summary table for each study (refer to COVID Serosurveillace dataset inventory)





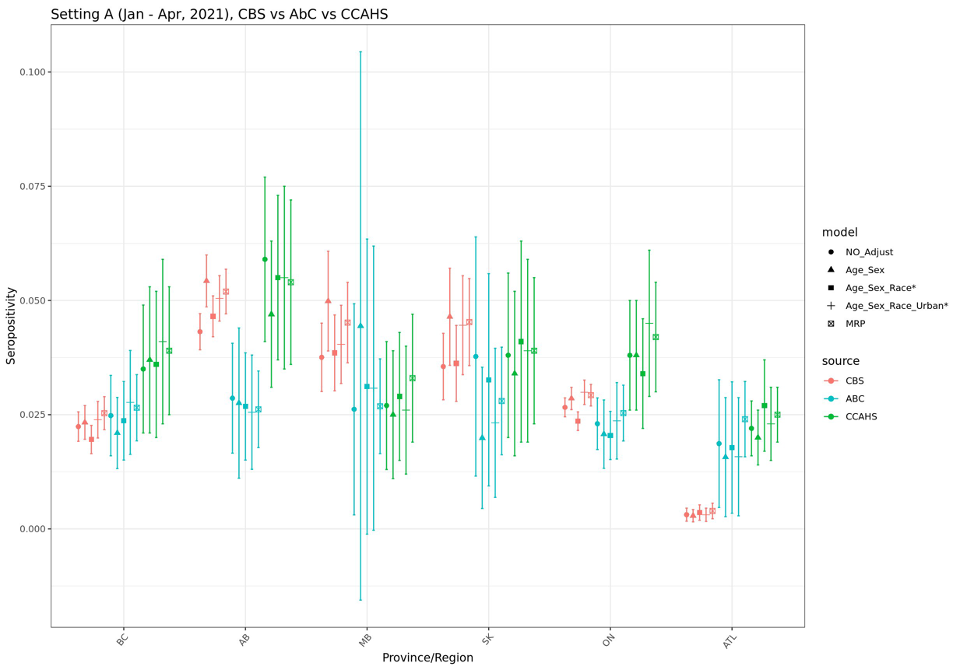
**Table S8:** Descriptive table for each time cuttings

**Table S9:** CV (Coefficient of Variance) for seropositivity estimations for each province/region for different time cuttings.

**Table S10:** CV (Coefficient of Variance) for seropositivity estimations for each province/region for different time cuttings.

##### 

# Supplemental figures



**Figure S1:** Setting A: Seropositivity During Vaccination Period (Jan – Apr, 2021, CBS, AbC, CCAHS I)

**Figure S2:** Setting A1: Seropositivity for Elder Group During Vaccination Period (Mar – Apr, 2021, CBS, AbC, CCAHS I, CLSA)

**Figure S3:** Setting A2: Seropositivity for Elder Group During Vaccination Period (Mar – Apr, 2021, CBS, AbC, CCAHS I, CLSA, CanPath)A graph with red and blue lines

Description automatically generated

**Figure S4:** Setting B: Seropositivity During Vaccination Period (Jul – Aug, 2021, CBS, AbC, CanPath)A graph of a graph

Description automatically generated with medium confidence

**Figure S5:** Setting C: Seropositivity During Omicron (Jan – Apr, 2022, CBS, AbC, CanPath)

**Figure S6:** Setting C1: Seropositivity During Omicron (Apr, 2022, CBS, AbC, CCAHS II, CanPath)

**Figure S7:** Setting D: Seropositivity for Elder Group During Vaccination Period (Mar – Aug, 2021, CBS, CLSA, CanPath)



**Figure S8:** Setting E: Seropositivity During Omicron (Apr – Sep, 2022, CBS, CCAHS II, CanPath)A graph with green and red lines

Description automatically generated

**Figure S9:** Setting F: Seropositivity During Vaccination Period (Dec 21 – Apr 2022, CBS, CCAHS I)A graph with different colored lines

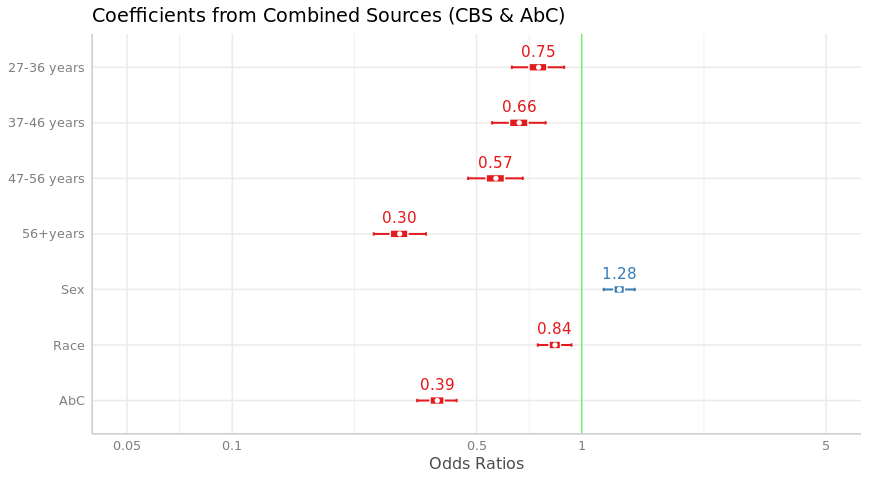
Description automatically generated

**Figure S10:** Setting G: Seropositivity for Alberta During Omicron (Jan – Apr, 2022, CBS, AbC, APL)A graph of different colored lines

Description automatically generated

**Figure S11:** Setting H: Seropositivity for Elder Group During Vaccination Period (Mar – Apr &Jul – Aug 2021, CBS, AbC, CLSA, CanPath)

**Figure S12:** Evaluate the effect for MRP for different simulated biased sample from the CBS



**Figure S13:** Odds Ratio Across combined dataset from multiple data sources